



(19)

Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

EP 0 897 980 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:  
24.02.1999 Bulletin 1999/08

(51) Int Cl<sup>6</sup>: C12N 15/12, C12N 15/11,  
C12N 15/85, C12N 5/10,  
C12Q 1/68, C07K 14/715,  
C07K 16/28, A61K 31/70,  
A61K 38/17, A61K 48/00,  
G01N 33/53

(21) Application number: 98306324.9

(22) Date of filing: 07.08.1998

(84) Designated Contracting States:  
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE  
Designated Extension States:  
AL LT LV MK RO SI

(72) Inventors:  
• Gupta, Shalley Kant  
King of Prussia, Pennsylvania 19406 (US)  
• Pillaiisettil, Kodandaram  
King of Prussia, Pennsylvania 19406 (US)

(30) Priority: 20.08.1997 US 56601 P  
24.07.1998 US

(74) Representative:  
Connell, Anthony Christopher et al  
SmithKline Beecham plc  
Corporate Intellectual Property,  
Two New Horizons Court  
Brentford, Middlesex TW8 9EP (GB)

(71) Applicant: SMITHKLINE BEECHAM  
CORPORATION  
Philadelphia Pennsylvania 19103 (US)

(54) CXCR4B: A human splice variant of CXCR4 chemokine receptor

(57) The CXCR4B polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed

are methods for utilising CXCR4B polypeptides and polynucleotides in therapy, and diagnostic assays for such.

EP 0 897 980 A2

**Description**

[0001] This application claims the benefit of U.S. Provisional Application No. 60/056,601, filed August 20, 1997, whose contents are herein incorporated by reference.

5

**Field of the Invention**

[0002] This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in therapy and in identifying compounds which may be agonists, antagonists and/or inhibitors which are 10 potentially useful in therapy, and to production of such polypeptides and polynucleotides.

**Background of the Invention**

[0003] The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach is rapidly superceding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

[0004] Functional genomics relies heavily on the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and 20 characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

[0005] It is well established that many medically significant biological processes are mediated by proteins participating in signal transduction pathways that involve G-proteins and/or second messengers, e.g., cAMP (Lefkowitz, *Nature*, 1991, 351:353-354). Herein these proteins are referred to as proteins participating in pathways with G-proteins or PPG proteins. Some examples of these proteins include the GPC receptors, such as those for adrenergic agents and 25 dopamine (Kobilka, B.K., et al., *Proc. Natl Acad. Sci., USA*, 1987, 84:46-50; Kobilka, B.K., et al., *Science*, 1987, 238: 650-656; Bunzow, J.R., et al., *Nature*, 1988, 336:783-787), G-proteins themselves, effector proteins, e.g., phospholipase C, adenyl cyclase, and phosphodiesterase, and actuator proteins, e.g., protein kinase A and protein kinase C (Simon, M.I., et al., *Science*, 1991, 252:802-8).

[0006] For example, in one form of signal transduction, the effect of hormone binding is activation of the enzyme, 30 adenylate cyclase, inside the cell. Enzyme activation by hormones is dependent on the presence of the nucleotide, GTP. GTP also influences hormone binding. A G-protein connects the hormone receptor to adenylate cyclase. G-protein was shown to exchange GTP for bound GDP when activated by a hormone receptor. The GTP-carrying form then binds to activated adenylate cyclase. Hydrolysis of GTP to GDP, catalyzed by the G-protein itself, returns the G-protein to its basal, inactive form. Thus, the G-protein serves a dual role, as an intermediate that relays the signal from 35 receptor to effector, and as a clock that controls the duration of the signal.

[0007] The membrane protein gene superfamily of G-protein coupled receptors has been characterized as having seven putative transmembrane domains. The domains are believed to represent transmembrane  $\alpha$ -helices connected by extracellular or cytoplasmic loops. G-protein coupled receptors include a wide range of biologically active receptors, such as hormone, viral, growth factor and neuroreceptors.

[0008] G-protein coupled receptors (otherwise known as 7TM receptors) have been characterized as including these 40 seven conserved hydrophobic stretches of about 20 to 30 amino acids, connecting at least eight divergent hydrophilic loops. The G-protein family of coupled receptors includes dopamine receptors which bind to neuroleptic drugs used for treating psychotic and neurological disorders. Other examples of members of this family include, but are not limited to, calcitonin, adrenergic, endothelin, cAMP, adenosine, muscarinic, acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating hormone, opsins, endothelial differentiation gene-1, rhodopsins, odorant, and cytomegalovirus 45 receptors.

[0009] Most G-protein coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. The 7 transmembrane regions are designated as TM1, TM2, TM3, TM4, TM5, TM6, and TM7. TM3 has been implicated in signal transduction.

[0010] Phosphorylation and lipidation (palmitylation or farnesylation) of cysteine residues can influence signal transduction of some G-protein coupled receptors. Most G-protein coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxy terminus. For several G-protein coupled receptors, such as the  $\beta$ -adrenoreceptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

[0011] For some receptors, the ligand binding sites of G-protein coupled receptors are believed to comprise hydrophilic sockets formed by several G-protein coupled receptor transmembrane domains, said sockets being surrounded by hydrophobic residues of the G-protein coupled receptors. The hydrophilic side of each G-protein coupled receptor transmembrane helix is postulated to face inward and form a polar ligand binding site. TM3 has been implicated in 55

several G-protein coupled receptors as having a ligand binding site, such as the TM3 aspartate residue. TM5 serines, a TM6 asparagine and TM6 or TM7 phenylalanines or tyrosines are also implicated in ligand binding.

[0012] G-protein coupled receptors can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters (see, Johnson et al., *Endoc. Rev.*, 1989, 10:317-331). Different G-protein  $\alpha$ -

5 subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G-protein coupled receptors has been identified as an important mechanism for the regulation of G-protein coupling of some G-protein coupled receptors. G-protein coupled receptors are found in numerous sites within a mammalian host. Over the past 15 years, nearly 350 therapeutic agents targeting 7 transmembrane (7 TM) receptors have been successfully introduced onto the market.

10

### Summary of the Invention

[0013] The present invention relates to CXCR4B, in particular CXCR4B polypeptides and CXCR4B polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using 15 such polypeptides and polynucleotides, including the treatment of infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes; obesity; anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; hypotension; hypertension; urinary retention; osteoporosis; angina pectoris; myocardial infarction; stroke; ulcers; asthma; allergies; benign prostatic hypertrophy; migraine; vomiting; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, depression, delirium, dementia, 20 and severe mental retardation; and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, hereinafter referred to as "the Diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with CXCR4B imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate CXCR4B activity or levels.

25

### Description of the Invention

[0014] In a first aspect, the present invention relates to CXCR4B polypeptides. Such peptides include isolated 30 polypeptides comprising an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include those comprising the amino acid of SEQ ID NO:2.

[0015] Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has 35 at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include the polypeptide of SEQ ID NO:2.

[0016] Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

[0017] Polypeptides of the present invention are believed to be members of the G-Protein Coupled Receptor family 40 of polypeptides. They are therefore of interest because G-Protein Coupled Receptors, more than any other gene family, have been the targets for pharmaceutical intervention. CXCR4 receptor has been characterized as one of the co-receptors for HIV infection. Additionally, CXCR4 is implicated in cardiac development by knockout studies in mice, and is also thought to be involved in inflammation, endothelial cell migration and vasculogenesis. The splice variant CXCR4B (SEQ ID NO:2) is unique at its N' terminal sequence. Ligand binding studies have shown a clear indication of 45 the importance of the N' terminal sequence in the function of various chemokine receptors. Studies with stably transfected RBL cells show that CXCR4B also uses SDF-1 as its ligand. This indicates that the CXCR4B splice form with its unique N' terminal sequence, expression pattern and function is involved in HIV infection, cardiac development, inflammation and embryogenesis. These properties are hereinafter referred to as "CXCR4B activity" or "CXCR4B polypeptide activity" or "biological activity of CXCR4B". Also included amongst these activities are antigenic and 50 immunogenic activities of said CXCR4B polypeptides, in particular the antigenic and immunogenic activities of the polypeptide of SEQ ID NO:2. Preferably, a polypeptide of the present invention exhibits at least one biological activity of CXCR4B.

[0018] The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an additional amino acid sequence which 55 contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

[0019] The present invention also includes variants of the aforementioned polypeptides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with

like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

5 [0020] Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

10 [0021] In a further aspect, the present invention relates to CXCR4B polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising 15 the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2.

20 [0022] Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

25 [0023] Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ ID NO: 1 over the entire length of SEQ ID NO: 1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO: 1 as well as the polynucleotide of SEQ ID NO: 1.

30 [0024] The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

35 [0025] The nucleotide sequence of SEQ ID NO:1 shows homology with CXCR4 (Herzog et.al., DNA Cell.Biol. 12, 465-471; 1993). The nucleotide sequence of SEQ ID NO: 1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 336 to 1403) encoding a polypeptide of 356 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO: 1 or it may be a sequence other than the one contained in SEQ ID NO: 1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of SEQ ID NO:2 is structurally related to other proteins of the G-Protein coupled receptor family, having homology and/or structural similarity with CXCR4 (Herzog et.al., DNA Cell.Biol. 12, 465-471 1993).

40 [0026] Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one CXCR4B activity.

45 [0027] The present invention also relates to partial or other polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO: 1 and SEQ ID NO:2.

50 [0028] Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide comprising:

55 (a) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

60 (b) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

65 (c) the polynucleotide of SEQ ID NO:3; or

70 (d) a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4;

75 as well as the polynucleotide of SEQ ID NO:3.

[0029] The present invention further provides for a polypeptide which:

80 (a) comprises an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more

preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(b) has an amino acid sequence which is at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

5 (c) comprises the amino acid of SEQ ID NO:4; and

(d) is the polypeptide of SEQ ID NO:4;

as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

10 [0030] The nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognized by those skilled in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al.*, *Nature* 377 (supp) 3, 1995). Accordingly, the nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded therefrom are therefore subject to the same inherent limitations in sequence accuracy. Furthermore, the peptide sequence encoded by SEQ ID NO:3 comprises a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

15 [0031] Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human pancreas, thymus and neutrophils, using the expressed sequence tag (EST) analysis (Adams, M.D., *et al.*, *Science* (1991) 252:1651-1656; Adams, M.D. *et al.*, *Nature*, (1992) 20 355:632-634; Adams, M.D., *et al.*, *Nature* (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

25 [0032] When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

30 [0033] Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

35 [0034] Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding homologs and orthologs from species other than human) that have a high sequence similarity to SEQ ID NO: 1. Typically these nucleotide sequences are 70% identical, 40 preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides.

45 [0035] A polynucleotide encoding a polypeptide of the present invention, including homologs and orthologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared 50 salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof.

55 [0036] The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during 1 st strand cDNA synthesis.

[0037] There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or

extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and

5 an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analyzed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

10 [0038] Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host 15 cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

15 [0039] For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be 20 effected by methods described in many standard laboratory manuals, such as Davis et al., *Basic Methods in Molecular Biology* (1986) and Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

25 [0040] Representative examples of appropriate hosts include bacterial cells, such as *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

30 [0041] A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. 35 Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL* (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic 40 space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

45 [0042] If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

50 [0043] Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

55 [0044] This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterized by the polynucleotide of SEQ ID NO: 1 which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

[0045] Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically

by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled CXCR4B nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (e.g., Myers *et al.*, *Science* (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, *Proc Natl Acad Sci USA* (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprising CXCR4B nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, *Science*, Vol 274, pp 610-613 (1996)).

**[0046]** The diagnostic assays offer a process for diagnosing or determining a susceptibility to the Diseases through detection of mutation in the CXCR4B gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

**[0047]** Thus in another aspect, the present invention relates to a diagnostic kit which comprises:

- 25 (a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof;
- (b) a nucleotide sequence complementary to that of (a);
- (c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or
- (d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

**[0048]** It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes; obesity; anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; hypotension; hypertension; urinary retention; osteoporosis; angina pectoris; myocardial infarction; stroke; ulcers; asthma; allergies; benign prostatic hypertrophy; migraine; vomiting; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, depression, delirium, dementia, and severe mental retardation; and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, amongst others.

**[0049]** The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, *Mendelian Inheritance in Man* (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

**[0050]** The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

**[0051]** The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

**[0052]** Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983)

4:72) and the EBV-hybridoma technique (Cole *et al.*, MONOCLONAL ANTIBODIES AND CANCER THERAPY, pp. 77-96, Alan R. Liss, Inc., 1985).

[0053] Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic 5 mice, or other organisms, including other mammals, may be used to express humanized antibodies.

[0054] The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

[0055] Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others.

10 [0056] In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be 15 cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

20 [0057] Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, 25 delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

[0058] A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is 30 preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening 35 agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

40 [0059] Polypeptides of the present invention are responsible for many biological functions, including many disease states, in particular the Diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such Diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for 45 example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan *et al.*, *Current Protocols in Immunology* 1(2):Chapter 5 (1991)).

50 [0060] The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the 55 agonist by the presence of the candidate compound is observed. Constitutively active polypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention,

to form a mixture, measuring CXCR4B activity in the mixture, and comparing the CXCR4B activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and CXCR4B polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, *J Mol Recognition*, 8:52-58 (1995); and K. Johanson *et al.*, *J Biol Chem*, 270(16):9459-9471 (1995)).

[0061] One screening technique includes the use of cells which express the receptors of this invention (for example, transfected CHO cells) in a system which measures extracellular pH or intracellular calcium changes caused by receptor activation. In this technique, compounds may be contacted with cells expressing a receptor polypeptide of the present invention. A second messenger response, e.g., signal transduction, pH changes, or changes in calcium level, is then measured to determine whether the potential compound activates or inhibits the receptor. Another method involves screening for receptor inhibitors by determining inhibition or stimulation of receptor-mediated cAMP and/or adenylate cyclase accumulation. Such a method involves transfecting a eukaryotic cell with a receptor of this invention to express the receptor on the cell surface. The cell is then exposed to potential antagonists in the presence of the receptor of this invention. The amount of cAMP accumulation is then measured. If the potential antagonist binds the receptor, and thus inhibits receptor binding, the levels of receptor-mediated cAMP, or adenylate cyclase, activity will be reduced or increased. Another method for detecting agonists or antagonists for the receptor of the present invention is the yeast based technology as described in U.S. Patent No. 5,482,835.

[0062] The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

[0063] The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance,  $^{125}$ I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

[0064] Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

[0065] Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

- 40 (a) a polypeptide of the present invention;
- (b) a recombinant cell expressing a polypeptide of the present invention;
- (c) a cell membrane expressing a polypeptide of the present invention; or
- (d) antibody to a polypeptide of the present invention;

45 which polypeptide is preferably that of SEQ ID NO:2.

[0066] It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

[0067] It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

- 50 (a) determining in the first instance the three-dimensional structure of the polypeptide;
- (b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;
- (c) synthesizing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and
- (d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors. It will be further appreciated that this will normally be an interactive process.

[0068] In a further aspect, the present invention provides methods of treating abnormal conditions such as, for in-

stance, infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2; pain; cancers; diabetes; obesity; anorexia; bulimia; asthma; Parkinson's disease; acute heart failure; hypotension; hypertension; urinary retention; osteoporosis; angina pectoris; myocardial infarction; stroke; ulcers; asthma; allergies; benign prostatic hypertrophy; migraine; vomiting; psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, depression, delirium, dementia, and severe mental retardation; and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome, related to either an excess of, or an under-expression of, CXCR4B polypeptide activity.

**[0069]** If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the CXCR4B polypeptide.

**[0070]** In still another approach, expression of the gene encoding endogenous CXCR4B polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered (see, for example, O'Connor, *J Neurochem* (1991) 56:560 in *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices with the gene can be supplied (see, for example, Lee *et al.*, *Nucleic Acids Res* (1979) 6:3073; Cooney *et al.*, *Science* (1988) 241:456; Dervan *et al.*, *Science* (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

**[0071]** For treating abnormal conditions related to an under-expression of CXCR4B and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of CXCR4B by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in *Human Molecular Genetics*, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

**[0072]** In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

**[0073]** The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

**[0074]** The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

**[0075]** Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities

often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

[0076] Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as GCC. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO: 1 and/or a polypeptide sequence encoded thereby.

[0077] The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

[0078] "Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

[0079] "Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

[0080] "Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single-and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine.

[0081] A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

[0082] "Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62).

[0083] "Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally,

differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

[0083] "Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heijne, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S. F. et al., *J. Molec. Biol.* 215: 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Atschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Atschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

[0084] Parameters for polypeptide sequence comparison include the following:

25 1) Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)  
 Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci. USA*. 89:10915-10919(1992)  
 Gap Penalty: 12  
 Gap Length Penalty: 4

30 A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

[0085] Parameters for polynucleotide comparison include the following:

35 1) Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)  
 Comparison matrix: matches = +10, mismatch = 0  
 Gap Penalty: 50  
 Gap Length Penalty: 3

40 Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

[0086] A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

45 [0087] (1) Polynucleotide embodiments further include an isolated polynucleotide comprising a polynucleotide sequence having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to the reference sequence of SEQ ID NO: 1, wherein said polynucleotide sequence may be identical to the reference sequence of SEQ ID NO: 1 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO: 1 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleotides in SEQ ID NO: 1, or:

55

$$n_n \leq x_n - (x_n \bullet y),$$

wherein  $n_n$  is the number of nucleotide alterations,  $x_n$  is the total number of nucleotides in SEQ ID NO:1,  $y$  is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and  $\cdot$  is the symbol for the multiplication operator, and wherein any non-integer product of  $x_n$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_n$ . Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

5 [0088] By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such 10 alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The 15 number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$n_n \leq x_n - (x_n \cdot y),$$

20 wherein  $n_n$  is the number of amino acid alterations,  $x_n$  is the total number of amino acids in SEQ ID NO:2,  $y$  is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc.,  $\cdot$  is the symbol for the multiplication operator, and wherein any non-integer product of  $x_n$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_n$ .

25 [0089] (2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 50,60, 70, 80, 85, 90, 95, 97 or 100% identity to a polypeptide reference sequence of SEQ ID NO:2, wherein said polypeptide sequence may be identical to the reference sequence of SEQ ID NO: 2 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative 30 substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid alterations is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

35.

$$n_a \leq x_a - (x_a \cdot y),$$

40 wherein  $n_a$  is the number of amino acid alterations,  $x_a$  is the total number of amino acids in SEQ ID NO:2,  $y$  is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and  $\cdot$  is the symbol for the multiplication operator, and wherein any non-integer product of  $x_a$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_a$ .

45 [0090] By way of example, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the 50 reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

55

$$n_a \leq x_a - (x_a \cdot y),$$

wherein  $n_a$  is the number of amino acid alterations,  $x_a$  is the total number of amino acids in SEQ ID NO:2,  $y$  is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and  $\cdot$  is the symbol for the multiplication operator, and wherein

any non-integer product of  $x_a$  and  $y$  is rounded down to the nearest integer prior to subtracting it from  $x_a$ .

[0091] "Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

[0092] All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

### Examples

#### Example 1: Mammalian Cell Expression

[0093] The receptors of the present invention are expressed in either human embryonic kidney 293 (HEK293) cells or adherent dhfr CHO cells. To maximize receptor expression, typically all 5' and 3' untranslated regions (UTRs) are removed from the receptor cDNA prior to insertion into a pCDN or pCDNA3 vector. The cells are transfected with individual receptor cDNAs by lipofection and selected in the presence of 400 mg/ml G418. After 3 weeks of selection, individual clones are picked and expanded for further analysis. HEK293 or CHO cells transfected with the vector alone (i.e., without the cDNA clone) serve as negative controls. To isolate cell lines stably expressing the individual receptors, about 24 clones are typically selected and analyzed by Northern blot analysis. Receptor mRNAs are generally detectable in about 50% of the G418-resistant clones analyzed.

#### Example 2 Ligand bank for binding and functional assays.

[0094] A bank of over 200 putative receptor ligands has been assembled for screening. The bank comprises: transmitters, hormones and chemokines known to act via a human seven transmembrane (7TM) receptor; naturally occurring compounds which may be putative agonists for a human 7TM receptor, non-mammalian, biologically active peptides for which a mammalian counterpart has not yet been identified; and compounds not found in nature, but which activate 7TM receptors with unknown natural ligands. This bank is used to initially screen the receptor for known ligands, using both functional (i.e. calcium, cAMP, microphysiometer, oocyte electrophysiology, etc, see below) as well as binding assays.

#### Example 3: Ligand Binding Assays

[0095] Ligand binding assays provide a direct method for ascertaining receptor pharmacology and are adaptable to a high throughput format. The purified ligand for a receptor is radiolabeled to high specific activity (50-2000 Ci/mmol) for binding studies. A determination is then made that the process of radiolabeling does not diminish the activity of the ligand towards its receptor. Assay conditions for buffers, ions, pH and other modulators such as nucleotides are optimized to establish a workable signal to noise ratio for both membrane and whole cell receptor sources. For these assays, specific receptor binding is defined as total associated radioactivity minus the radioactivity measured in the presence of an excess of unlabeled competing ligand. Where possible, more than one competing ligand is used to define residual nonspecific binding.

#### Example 4: Functional Assay in Xenopus Oocytes

[0096] Capped RNA transcripts from linearized plasmid templates encoding the receptor cDNAs of the invention are synthesized in vitro with RNA polymerases in accordance with standard procedures. In vitro transcripts are suspended in water at a final concentration of 0.2 mg/ml. Ovarian lobes are removed from adult female toads, Stage V defolliculated oocytes are obtained, and RNA transcripts (10 ng/oocyte) are injected in a 50 nl bolus using a microinjection apparatus. Two electrode voltage clamps are used to measure the currents from individual *Xenopus* oocytes in response to agonist exposure. Recordings are made in Ca<sup>2+</sup> free Barth's medium at room temperature. The *Xenopus* system can be used to screen known ligands and tissue/cell extracts for activating ligands.

#### Example 5: Microphysiometric Assays

[0097] Activation of a wide variety of secondary messenger systems results in extrusion of small amounts of acid

from a cell. The acid formed is largely as a result of the increased metabolic activity required to fuel the intracellular signaling process. The pH changes in the media surrounding the cell are very small but are detectable by the CYTO-SENSOR microphysiometer (Molecular Devices Ltd., Menlo Park, CA). The CYTOSENSOR is thus capable of detecting the activation of a receptor which is coupled to an energy utilizing intracellular signaling pathway such as the G-protein coupled receptor of the present invention.

5 Example 6: Extract/Cell Supernatant Screening

10 [0098] A large number of mammalian receptors exist for which there remains, as yet, no cognate activating ligand (agonist). Thus, active ligands for these receptors may not be included within the ligands banks as identified to date. Accordingly, the 7TM receptor of the invention is also functionally screened (using calcium, cAMP, microphysiometer, oocyte electrophysiology, etc., functional screens) against tissue extracts to identify natural ligands. Extracts that produce positive functional responses can be sequentially subfractionated until an activating ligand is isolated and identified.

15 Example 8: Calcium and cAMP Functional Assays

20 [0099] 7TM receptors which are expressed in HEK 293 cells have been shown to be coupled functionally to activation of PLC and calcium mobilization and/or cAMP stimulation or inhibition. Basal calcium levels in the HEK 293 cells in receptor-transfected or vector control cells were observed to be in the normal, 100 nM to 200 nM, range. HEK 293 cells expressing recombinant receptors are loaded with fura 2 and in a single day > 150 selected ligands or tissue/cell extracts are evaluated for agonist induced calcium mobilization. Similarly, HEK 293 cells expressing recombinant receptors are evaluated for the stimulation or inhibition of cAMP production using standard cAMP quantitation assays. Agonists presenting a calcium transient or cAMP fluctuation are tested in vector control cells to determine if the response 25 is unique to the transfected cells expressing receptor.

30

35

40

45

50

55

## SEQUENCE INFORMATION

SEQ ID NO:1

5           1       GGC ACGAGGAGAG AGAGAACTAG TCTCGCGTTT TTTTTTCTTC

10           44      CCTCTAGTGG GCGGGGCAGA GGAGTTAGCC AAGATGTGAC TTTGAAACCC

15           94      TCAGCGTCTC AGTGCCCTTT TGTTCTAAC AAAGAATTAA GTAATTGGTT

20           144     CTACCAAAGA AGGATATAAT GAAGTCACTA TGGGAAAAGA TGGGGAGGAG

25           194     AGTTGTAGGA TTCTACATTA ATTCTCTTGT GCCCTTAGCC CACTACTTCA

30           244     GAATTTCCCTG AAGAAAGCAA GCCTGAATTG GTTTTTTAAA TTGCTTTAAA

35           294     AATTTTTTTT AACTGGGTTA ATGCTTGCTG AATTGGAAGT GAATGTCCAT

40           344     TCCTTTGCCT CTTTGCAGA TATAACTTC AGATAACTAC ACCGAGGAAA

45           394     TGGGCTCAGG GGACTATGAC TCCATGAAGG AACCTGTTT CCGTGAAGAA

50           444     AATGCTAATT TCAATAAAAT CTCCTGCC ACCATCTACT CCATCATCTT

55           494     CTTAAC TGCG ATTGTGGCA ATGGATTGGT CATCCTGGTC ATGGGTTACC

60           544     AGAAGAAACT GAGAAGCATG ACGGACAAGT ACAGGCTGCA CCTGTCAGTG

65           594     GCCGACCTCC TCTTGTCA CACGCTTCCC TTCTGGGCAG TTGATGCCGT

70           644     GCCAAACTGG TACTTTGGGA ACTTCCTATG CAAGGCAGTC CATGTCATCT

75           694     ACACAGTCAA CCTCTACAGC ACTGTCTCA TCCTGGCCTT CATCAGTCTG

80           744     GACCGCTACC TGGCCATCGT CCACGCCACC AACAGTCAGA GGCCAAGGAA

85           794     GCTGTTGGCT GAAAAGGTGG TCTATGTTGG CGTCTGGATC CCTGCCCTCC

90           844     TGCTGACTAT TCCCGACTTC ATCTTTGCCA ACGTCAGTGA GGCAGATGAC

95           894     AGATATATCT GTGACCGCTT CTACCCCAAT GACTTGTGGG TGGTTGTGTT

100           944     CCAGTTTCAG CACATCATGG TTGGCCTTAT CCTGCCTGGT ATTGTCATCC

105           994     TGTCTGCTA TTGCATTATC ATCTCCAAGC TGTCACACTC CAAGGGCCAC

1044 CAGAAGCGCA AGGCCCTCAA GACCACAGTC ATCCTCATCC TGGCTTCCTT  
 5 1094 CGCCTGTTGG CTGCCTTAAC ATTGGGAT CAGCATCGAC TCCTTCATCC  
 1144 TCCTGGAAAT CATCAAGCAA GGGTGTGAGT TTGAGAACAC TGTGCACAAG  
 10 1194 TGGATTCCA TCACCGAGGC CCTAGCTTC TTCCACTGTT GTCTGAACCC  
 15 1244 CATCCTCTAT GCTTCCTTG GAGCAAATT TAAAACCTCT GCCCAGCACG  
 1294 CACTCACCTC TGTGAGCAGA GGGTCCAGCC TCAAGATCCT CTCCAAAGGA  
 1344 AAGCGAGGTG GACATTCAATC TGTTCCACT GAGTCTGAGT CTTCAAGTTT  
 20 1394 TCACTCCAGC **TAACACAGAT** GTAAAAGACT TTTTTTATA CGATAAATAA  
 1444 CTTTTTTTTA AGTTACACAT TTTTCAGATA TAAAAGACTG ACCAATATTG  
 25 1494 TACAGTTTTT ATTGCTTGTT GGATTTTGT CTTGTGTTTC TTTAGTTTT  
 1544 GTGAAGTTA ATTGACTTAT TTATATAAAAT TTTTTTGTT TCATATTGAT  
 1594 GTGTGTCTAG GCAGGACCTG TGGCCAAGTT CTTAGTTGCT GTATGTCTCG  
 30 1644 TGGTAGGACT GTAGAAAAGG GAACTGAACA TTCCAGAGCG TGTAGTGAAT  
 1694 CACGTAAAGC TAGAAATGAT CCCCAGCTGT TTATGCATAG ATAATCTCTC  
 1744 CATTCCCGTG GAACGTTTTT CCTGTTCTTA AGACGTGATT TTGCTGTAGA  
 40 1794 AGATGGCACT TATAACCAAA GCCCAAAGTG GTATAGAAAT GCTGGTTTT  
 1844 CAGTTTCAG GAGTGGGTTG ATTCAGCAC CTACAGTGTA CAAGTCTTGK  
 1894 ATTAAGTTGK TAATAAAAGT ACATGTTAAA CTTAAAAAAA AAAAAAAA  
 50 1944 A

The start codon (ATG at position 336nt) and the stop codon (TAA at position 1406) is highlighted in italics. The splice acceptor site is also highlighted.

## SEQ ID NO:2

1 *MSIPLPLLQI* YTSDNYTEEM GSGDYDSMKE PCFREENANF NKIFLPTIYS  
 5  
 51 IIFLTGIVGN GLVILVMGYQ KKLRSMTDKY RLHLSVADLL FVITLPFWAV  
 10 101 DAVANWYFGN FLCKAVHVIY TVNLYSSVLI LAFISLDRYL AIVHATNSQR  
 101  
 151 PRKLLAEKVV YVGWVIPALL LTIPDFIFAN VSEADDRYIC DRFYPNDLWV  
 15 201 VVFQFQHIMV GLILPGIVIL SCYCIISKL SHSKGHQKRK ALKTTVILIL  
 201  
 251 AFFACWLWYY *IGISIDSFIL* LEIIKQGCEF ENTVHKWISI TEALAFFHCC  
 20 301 LNPILYAFLG AKFKTSQAQHA LTSVSRGSSL KILSKGKRGG HSSVSTESES  
 301  
 351 SSFHSS\*

The unique N'terminal extracellular domain generated by alternate splicing mechanism is highlighted in italics.

25

## SEQ ID NO:3

ggcagagggagagagagaactagtctcgcttttcttctccctctagtgccggggcagaggagtttagc  
 30 caagatgtgactttgaaaccctcagcgtctcagtgcctttgttctaaacaaagaattttgttaattggttc  
 taccaaagaaggatataatgaagtcactatggaaaagatggggaggagagtgttaggattctacattaatt  
 ctcttgcccttagcccactacttcagaatttcctgaagaaagcaagcctgaattggttttaaattgtct  
 ttaaaaatttttaactgggtaatgcttgcgtgaatttggaaagtgaATGTCCATTCCCTTGCCTCTTTGCT  
 35 AGATATACACTTCAGATAACTACACCGAGGAAATGGCTCAGGGACTATGACTCCATGAAGGAACCTGT  
 TTCCGTGAAGAAAATGCTAATTCAATAAAATCTCCTGCCACCCTACTCCATCATCTTCTTAACGGC  
 ATTGTGGCAATGGATTGGTCATCCTGGTCATGGGTTACCAAGAAAATGAGAAGCATGACGGACAAGTAC  
 AGGCTGCACCTGTCAGTGGCCGACCTCCTTTGTC  
 40

## SEQ ID NO:4

45 *MSIPLPLLQI* YTSDNYTEEMGSGDYDSMKEPCFREENANFNKIFLPTIYSI IIFLTGIVGNGLVILVMGYQKK  
 LRSMTDKYRLHLSVADLLFV

50

55

Annex to the description

[0100]

5

## SEQUENCE LISTING

## (1) GENERAL INFORMATION

## (i) APPLICANT

10 (A) NAME: SMITHKLINE BEECHAM CORPORATION  
 (B) STREET: ONE FRANKLIN PLAZA  
 (C) CITY: PHILADELPHIA  
 (D) STATE OR PROVINCE: PENNSYLVANIA  
 (E) COUNTRY: USA  
 (F) POSTAL CODE: 19103

15 (ii) TITLE OF THE INVENTION: CXCR4B: A HUMAN SPLICE VARIANT  
 OF CXCR4 CHEMOKINE RECEPTOR

(iii) NUMBER OF SEQUENCES: 4

## (iv) COMPUTER-READABLE FORM:

20 (A) MEDIUM TYPE: Diskette  
 (B) COMPUTER: IBM Compatible  
 (C) OPERATING SYSTEM: DOS  
 (D) SOFTWARE: FastSEQ for Windows Version 2.0

## (v) CURRENT APPLICATION DATA:

25 (A) APPLICATION NUMBER: TO BE ASSIGNED

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 1944 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

40	GGCACGAGGA GAGAGAGAAC TAGTCTCGCG TTTTTTTTC TTCCCTCTAG TGGGCGGGGC AGAGGAGTTA GCCAAGATGT GACTTGAAA CCCTCAGCGT CTCAGTGCC TTTGTTCTA	60 120
	AAACAAAGAAT TTTGTAATTG GTTCTACCAA AGAAGGATAT AATGAAGTC CAATGGAAA AGATGGGGAG GAGAGTTGTA GGATTCTACA TTAATTCTCT TGTGCCCTTA GCCCACTACT	180 240
	TCAGAATTTC CTGAAGAACG CAAGCCTGAA TTGGTTTTT AAATTGCTTT AAAAATTTT TTTAACTGGG TTAATGCTG CTGAATTGGA AGTGAATGTC CATTCCTTTG CCTCTTTG	300 360
	AGATATACAC TTCAGATAAC TACACCGAGG AAATGGGCTC AGGGGACTAT GACTCCATGA AGGAACCCCTG TTTCCGTGAA GAAAATGCTA ATTCAATAA AATCTCCCTG CCCACCATCT	420 480
45	ACTCCATCAT CTTCTTAACG GGCATTTGTT GCAATGGATT GTCATCCTG GTCATGGTT ACCAGAAGAA ACTGAGAACG ATGACGGACA AGTACAGGCT GCACCTGTCA GTGCCGACCC	540 600
	TCCTCTTGT CATCACGCTT CCCTCTGGG CAGTTGATGC CGTGGCAAAC TGGTACTTTG GGAACCTCCT ATGCAAGGCA GTCCATGTCA TCTACACAGT CAACCTCTAC AGCAGTGTCC	660 720
	TCATCCTGGC CTTCATCAGT CTGGACCGCT ACCTGGCCAT CGTCCACGCC ACCAACAGTC AGAGGCCAAG GAAGCTGTG TGCTGAAAAGG TGGTCTATGT TGGCGTCTGG ATCCCTGCC	780 840
50	TCCTGCTGAC TATTCCGAC TTCATCTTG CCAACGTCA TGAGGCAGAT GACAGATATA TCTGTGACCG CTTCTACCCC AATGACTTGT GGGTGGTTGT GTTCCAGTTT CAGCACATCA	900 960
	TGGTGGCCT TATCCTGCTT GGTATTGTC TCCCTGCTCTG CTATTGCAATT ATCATCTCCA AGCTGTACA CTCCAAGGGC CACCAGAACG GCAAGGCCCT CAAGACCACA GTCATCCTCA	1020 1080
	TCCCTGGCTT CTTCGCTGT TGGCTGCCTT ACTACATTGG GATCAGCCTC GACTCCTTCA TCCCTCTGGA AATCATCAAG CAAGGGTGTG AGTTTGAGAA CACTGTGCAC AAGTGGATT	1140 1200
55	CCATCACCGA GGCCCTAGCT TTCTTCCACT GTTGTGCGA CCCCACATCTC TATGCTTCC TTGGAGCCAA ATTTAAACAG TCTGCCAGC ACGCACTCAC CTCTGTGAGC AGAGGGTCCA	1260 1320
	GCCTCAAGAT CCTCTCCAAA GGAAAGCGAG GTGGACATTC ACTGTGTTCC ACTGAGTCTG	1380

5	AGTCTTCAAG	TTTTCACTCC	AGCTAACACA	GATGTAAG	ACTTTTTT	ATACGATAAA	1440
	TAACCTTTTT	TTAAGTTACA	CATTTTCAG	ATATAAAAGA	CTGACCAATA	TTGTACAGTT	1500
	TTTATTGCTT	GTTGGATTTT	TGTCTGTGT	TTCTTAGTT	TTTGTGAAGT	TTAATGACT	1560
	TATTTATATA	AATTTTTTTT	GTTCATATT	GATGTGTGTC	TAGGCAGGAC	CTGTGCCAA	1620
	GTTCCTTAGTT	GCTGTATGTC	TCGTGGTAGG	ACTGTAGAAA	AGGAACTGTA	ACATTCCAGA	1680
	GCGTGTAGTG	AATCACGTA	AGCTAGAAAT	GATCCCCAGC	TGTTTATGCA	TAGATAATCT	1740
	CTCCATTCCC	GTGGAACGTT	TTCCCTGTTC	TTAAGACGTC	ATTTTGCTGT	AGAAGATGGC	1800
	ACTTATAACC	AAAGCCCCAA	GTGGTATAGA	AATGCTGGTT	TTTCAGTTT	CAGGAGTGGG	1860
10	TTGATTCAG	CACCTACAGT	GTACAAGTCT	TGKATTAAGT	TGKTAATAAA	AGTACATGTT	1920
	AAACTTAAAA	AAAAAAA	AAAAA				1944

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 356 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

20	Met	Ser	Ile	Pro	Leu	Pro	Leu	Leu	Gln	Ile	Tyr	Thr	Ser	Asp	Asn	Tyr	
	1			5					10				15				
25	Thr	Glu	Glu	Met	Gly	Ser	Gly	Asp	Tyr	Asp	Ser	Met	Lys	Glu	Pro	Cys	
				20					25				30				
30	Phe	Arg	Glu	Glu	Asn	Ala	Asn	Phe	Asn	Lys	Ile	Phe	Leu	Pro	Thr	Ile	
					35				40			45					
35	Tyr	Ser	Ile	Ile	Phe	Leu	Thr	Gly	Ile	Val	Gly	Asn	Gly	Leu	Val	Ile	
					50				55			60					
40	Leu	Val	Met	Gly	Tyr	Gln	Lys	Lys	Leu	Arg	Ser	Met	Thr	Asp	Lys	Tyr	
					65				70			75			80		
45	Arg	Leu	His	Leu	Ser	Val	Ala	Asp	Leu	Leu	Phe	Val	Ile	Thr	Leu	Pro	
					85				90			95					
50	Phe	Trp	Ala	Val	Asp	Ala	Val	Ala	Asn	Trp	Tyr	Phe	Gly	Asn	Phe	Leu	
					100				105			110					
55	Cys	Lys	Aia	Val	His	Val	Ile	Tyr	Thr	Val	Asn	Leu	Tyr	Ser	Ser	Val	
					115				120			125					
60	Leu	Ile	Leu	Ala	Phe	Ile	Ser	Leu	Asp	Arg	Tyr	Leu	Ala	Ile	Val	His	
					130				135			140					
65	Ala	Thr	Asn	Ser	Gln	Arg	Pro	Arg	Lys	Leu	Leu	Ala	Glu	Lys	Val	Val	
					145				150			155			160		
70	Tyr	Val	Gly	Val	Trp	Ile	Pro	Ala	Leu	Leu	Leu	Thr	Ile	Pro	Asp	Phe	
					165				170			175					
75	Ile	Phe	Aia	Asn	Val	Ser	Glu	Ala	Asp	Asp	Arg	Tyr	Ile	Cys	Asp	Arg	
					180				185			190					
80	Phe	Tyr	Pro	Asn	Asp	Leu	Trp	Val	Val	Val	Phe	Gln	Phe	Gln	His	Ile	
					195				200			205					
85	Met	Val	Gly	Ile	Ile	Leu	Leu	Pro	Gly	Ile	Val	Ile	Leu	Ser	Cys	Tyr	Cys
					210				215			220					
90	Ile	Ile	Ile	Ser	Lys	Leu	Ser	His	Ser	Lys	Gly	His	Gln	Lys	Arg	Lys	
					225				230			235			240		
95	Ala	Leu	Lys	Thr	Thr	Val	Ile	Leu	Ile	Leu	Ala	Phe	Phe	Ala	Cys	Trp	
					245				250			255					
100	Leu	Pro	Tyr	Tyr	Ile	Gly	Ile	Ser	Ile	Asp	Ser	Phe	Ile	Leu	Leu	Glu	
					260				265			270					
105	Ile	Ile	Lys	Gln	Gly	Cys	Glu	Phe	Glu	Asn	Thr	Val	His	Lys	Trp	Ile	
					275				280			285					
110	Ser	Ile	Thr	Glu	Ala	Leu	Ala	Phe	Phe	His	Cys	Cys	Ile	Asn	Pro	Ile	
					290				295			300					
115	Leu	Tyr	Aia	Phe	Leu	Gly	Ala	Lys	Phe	Lys	Thr	Ser	Ala	Gln	His	Ala	
					305				310			315			320		
120	Leu	Thr	Ser	Val	Ser	Arg	Gly	Ser	Ser	Leu	Lys	Ile	Leu	Ser	Lys	Gly	

	325	330	335
	Lys Arg Gly Gly His Ser Ser Val Ser Thr Glu Ser Glu Ser Ser Ser		
	340	345	350
5	Phe His Ser Ser		
	355		

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 611 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

20	GGCAGAGGGGA GAGAGAGAAC TAGTCTCGCG TTTTCTTTC TTCCCTCTAG TGGGCGGGGC	60
	AGAGGAGTTA GCCAAGATGT GACTTTGAAA CCCTCAGCGT CTCAGTGCC TTTGTTCTA	120
	AACAAAGAAT TTTGTAATTG GTTCTACCAA AGAAGGATAT AATGAAGTCA CTATGGGAAA	180
	AGATGGGGAG GAGAGTTGTA GGATTCTACA TTAATTCTCT TGTGCCCTTA GCCCACTACT	240
	TCAGAATTTC CTGAAGAAAG CAAGCTGAA TTGGTTTTT AAATTGCTTT AAAAATTTT	300
	TTTAACTGGG TTAATGCTT CTGAATTGGA AGTGAATGTC CATTCCCTTG CCTCTTTGC	360
25	AGATATACAC TTCAGATAAC TACACCGAGG AAATGGGCTC AGGGGACTAT GACTCCATGA	420
	AGGAACCTG TTTCCGTGAA GAAAATGCTA ATTCAATAA AATCTTCCCTG CCCACCATCT	480
	ACTCCATCAT CTTCTTAACT GGCATTGTGG GCAATGGATT GGTCATCCCTG GTCATGGTT	540
	ACCAGAAGAA ACTGAGAAGC ATGACGGACA AGTACAGGCT GCACCTGTCA GTGGCCGACC	600
	TCCTCTTGT C	611

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

40	Met Ser Ile Pro Leu Pro Leu Leu Gln Ile Tyr Thr Ser Asp Asn Tyr	
	1 5 10 15	
	Thr Glu Glu Met Gly Ser Gly Asp Tyr Asp Ser Met Lys Glu Pro Cys	
	20 25 30	
	Phe Arg Glu Glu Asn Ala Asn Phe Asn Lys Ile Phe Leu Pro Thr Ile	
	35 40 45	
45	Tyr Ser Ile Ile Phe Leu Thr Gly Ile Val Gly Asn Gly Leu Val Ile	
	50 55 60	
	Leu Val Met Gly Tyr Gln Lys Lys Leu Arg Ser Met Thr Asp Lys Tyr	
	65 70 75 80	
	Arg Leu His Leu Ser Val Ala Asp Leu Leu Phe Val	
50	85 90	

## Claims

55 1. An isolated polypeptide selected from the group consisting of:

(i) an isolated polypeptide comprising an amino acid sequence selected from the group having at least:

- (a) 70% identity;
- (b) 80% identity;
- (c) 90% identity; or
- (d) 95% identity

5 to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2;

- (ii) an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:2 or
- (iii) an isolated polypeptide which is the amino acid sequence of SEQ ID NO:2.

10 2. An isolated polynucleotide selected from the group consisting of:

- (i) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least

- (a) 70% identity;
- (b) 80% identity;
- (c) 90% identity; or
- (d) 95% identity;

15 to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2;

20 (ii) an isolated polynucleotide comprising a nucleotide sequence that has at least:

- (a) 70% identity
- (b) 80% identity;
- (c) 90% identity; or
- (d) 95% identity;

25 over its entire length to a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;

- (iii) an isolated polynucleotide comprising a nucleotide sequence which has at least:

30 (a) 70% identity;

- (b) 80% identity;
- (c) 90% identity; or
- (d) 95% identity;

35 to that of SEQ ID NO: 1 over the entire length of SEQ ID NO:1;

- (iv) an isolated polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;
- (vi) an isolated polynucleotide which is the polynucleotide of SEQ ID NO: 1; or (vi) an isolated polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a labelled probe having the sequence of SEQ ID NO: 1 or a fragment thereof.;

40 or a nucleotide sequence complementary to said isolated polynucleotide.

3. An antibody immunospecific for the polypeptide of claim 1.

4. A method for the treatment of a subject:

45 (i) in need of enhanced activity or expression of the polypeptide of claim 1 comprising:

- (a) administering to the subject a therapeutically effective amount of an agonist to said polypeptide; and/or
- (b) providing to the subject an isolated polynucleotide comprising a nucleotide sequence encoding said polypeptide in a form so as to effect production of said polypeptide activity *in vivo*.; or

50 (ii) having need to inhibit activity or expression of the polypeptide of claim 1 comprising:

- (a) administering to the subject a therapeutically effective amount of an antagonist to said polypeptide; and/or
- (b) administering to the subject a nucleic acid molecule that inhibits the expression of a nucleotide sequence encoding said polypeptide; and/or
- (c) administering to the subject a therapeutically effective amount of a polypeptide that competes with

said polypeptide for its ligand, substrate, or receptor.

5. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of the polypeptide of claim 1 in a subject comprising:

(a) determining the presence or absence of a mutation in the nucleotide sequence encoding said polypeptide in the genome of said subject; and/or  
 (b) analyzing for the presence or amount of said polypeptide expression in a sample derived from said subject.

10. 6. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of claim 1 which comprises a method selected from the group consisting of:

(a) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presence of a labeled competitor;

(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

25. 7. An agonist or an antagonist of the polypeptide of claim 1.

8. An expression system comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression system is present in a compatible host cell.

30. 9. A process for producing a recombinant host cell comprising transforming or transfecting a cell with the expression system of claim 8 such that the host cell, under appropriate culture conditions, produces a polypeptide comprising an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2.

35. 10. A recombinant host cell produced by the process of claim 9.

11. A membrane of a recombinant host cell of claim 10 expressing a polypeptide comprising an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2.

40. 12. A process for producing a polypeptide comprising culturing a host cell of claim 10 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture.

13. An isolated polynucleotide selected from the group consisting of:

45. (a) an isolated polynucleotide comprising a nucleotide sequence which has at least 70%, 80%, 90%, 95%, 97% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;  
 (b) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:3;  
 (c) the polynucleotide of SEQ ID NO:3; or  
 (d) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide which has at least 70%, 80%, 90%, 95%, 97-99% identity to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4.

50. 14. A polypeptide selected from the group consisting of:

55. (a) a polypeptide which comprises an amino acid sequence which has at least 70%, 80%, 90%, 95%, 97-99% identity to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;  
 (b) a polypeptide which has an amino acid sequence which is at least 70%, 80%, 90%, 95%, 97-99% identity to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

**EP 0 897 980 A2**

- (c) a polypeptide which comprises the amino acid of SEQ ID NO:4;
- (d) a polypeptide which is the polypeptide of SEQ ID NO:4;
- (e) a polypeptide which is encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

5

10

15

20



25

30

35

40

45

50

55